

# Statistical modeling of language universals

BALTHASAR BICKEL

## 1. Introduction

A particularly interesting type of universal is diachronic in nature: principled preferences that affect how languages change over time. Yet it is still unclear what kinds of methods are most suitable to evaluating universals of this kind when – as is normally the case – we cannot directly sample actual cases of language change. Although they do not embed their research into this discussion, Dunn et al. (2011) propose one specific method for assessing diachronic preferences. The method they propose has a potential of becoming a welcome addition to our toolkit. However, at least in its published form, the method has one immediate shortcoming: it does not allow identification of other factors that compete with universal principles – notably factors of areal histories.

In the following I argue that this is a critical shortcoming (Section 2), and I propose that there are other methods available that allow simultaneous estimation of the relevant impact of universal, areal, and any other factors on language change (Section 3). Section 4 reports on a case study and in the concluding section I summarize what I take to be the state of the art, identifying what is needed for further progress (Section 5).

## 2. What shapes linguistic distributions?

The distribution of linguistic structures (features, categories, constructions, etc.) can reflect at least two PRINCIPLED ways which determine whether and how languages change over time – in addition to any non-principled (random and idiosyncratic) fluctuation:<sup>1</sup>

---

1. What follows assumes that current distributions of typological structures are not just faithful copies of ancestral distributions at the time when language first emerged. Bickel 2011 provides statistical evidence for this assumption (*pace* Maslova 2000).

- (i) **MATCH:** We like best what suits us best. The **MATCH** factor covers the large class of cognitive/physiological and social/communicative principles that can be hypothesized to universally bias the evolution of specific structures in certain ways, e.g., the (hypothesized) preference for agent-before-patient orders or for word order harmony across phrase types. **MATCH** factors can manifest themselves through defining universally preferred pathways of change (e.g., natural pathways of sound change or common grammaticalization channels) or through universally preferred selection of variants (e.g., expansion of preferred constructions at the expense of others).
- (ii) **SPREAD:** We like best what is most popular around us. The **SPREAD** factor refers to those processes of areal diffusion, from borrowing to substrate effects, in which structures are copied just because they are present in a contact language (and often perceived as prestigious), and not for any intrinsic and area-independent advantage of the structures themselves. The effects of **SPREAD** factors are contingent on the contact history of languages and therefore tied to specific locations and historical events.

The difference between **MATCH** and **SPREAD** factors often coincides with the distinction between vertical and horizontal transmission in evolutionary biology, but it is by no means equivalent to it. Horizontal transmission, i.e., areal diffusion, is not only subject to **SPREAD** factors but also to **MATCH** factors: structures that are easier to process for example can be expected to spread in contact more easily than more difficult structures. Also, given certain types of contact situations, the diffusion of specific structures across languages may be motivated by the complexities of bilingual processing. In both these cases, areal diffusion reflects **MATCH** factors that have universal validity:<sup>2</sup> with a certain probability, the diffusion is expected to apply between all languages in contact, or between all languages in contact that meet certain structural or sociological conditions.

**SPREAD** factors, by contrast, apply only to specific locations at specific times, e.g., the spread of *have*-perfects in Europe after the decline of the Roman empire (Haspelmath 1998). Therefore, **SPREAD** factors can only be modeled via the geographical areas that they create, e.g., by assessing whether Europe is statistically different from the rest of Eurasia and the rest of the world with regard

---

2. So why talk about “match” factors and not simply “universals”? Because “universals” also include absolute universals, i.e., boundary conditions on what can be described as a human language, given our metalanguage; see Bickel 2010 for some discussion. Also, it is important to emphasize that universals are interesting only to the extent that they have a solid theory of match behind them, i.e., evidence for why a certain structure suits us better than other structures. Statistical distributions are only one piece of evidence for universals, and universals need more supporting evidence from other fields, such as psychology and anthropology, than they tended to receive in the past.

to the structures of interest. SPREAD factors cannot be adequately modeled via universal probabilities of horizontal transmission (i.e., borrowability estimates or “diffusion potentials”): a given universal probability for, say, *have*-perfects to be copied in language contact would not model a SPREAD factor but a MATCH factor based on some universal theory of why we expect some structures to universally diffuse more easily than others. All that matters for modeling a SPREAD factor here is the historical accidents in Europe that led to an increased frequency of *have*-perfects in this area. Generalizations beyond this make no sense for SPREAD factors.

Given the relevance of both MATCH and SPREAD factors, it follows that if one wants to establish the presence of MATCH effects, one not only needs to show that they have a principled, non-random effect on distributions, but one also needs to systematically control for SPREAD effects: for example, if one wants to defend the preference for agent-before-patient orders as a MATCH effect, one needs to show that the observed distributions (i) are significantly affected by non-random processes and (ii) hold independently of SPREADS that one might hypothesize based on what is known about language contact and diffusion. (And, conversely, if one wants to demonstrate SPREAD effects, one needs to control for MATCH effects.)

Following up on the tradition established by Dryer (1989), I submit that any method for testing MATCH effects can and should be evaluated in terms of how well it incorporates a control of SPREAD factors (and vice versa). At least in its published form, the method proposed by Dunn et al. 2011 does not incorporate such a control (while the method is arguably robust against areal diffusion that results from MATCH factors defining universal probabilities for borrowing: see Curie et al. 2009). So, if the method would establish that two structures are associated (“coupled”) in the majority<sup>3</sup> of families (“lineages”) of the world, we still wouldn’t know whether this results from MATCH effects. It could just as well result from SPREAD effects inflating the number of families that happen to have the specific association in some specific areas of the world. Conversely, if the method does not suggest a universal association, this does not necessarily rule out the presence of a MATCH effect: it is possible that the effect becomes statistically visible only in a model where SPREAD factors already account for some of the variation so that the MATCH factors can account for the rest – just like when in an experiment, a certain effect may become visible only after weeding out signal noise from other sources. The general point is that unless we

---

3. In the LINGTYP discussion (April 14, 2011), Michael Dunn suggested that he is interested in “couplings” in all or no families, not just in the majority. But this would be a non-statistical understanding of statistical universals that I will disregard as obsolete. Languages and families are not a finite universe and so all quantitative statements are necessarily statistical and here, counts of “none” or “all” do not have a privileged status (as Cysouw 2002 emphasized).

can discriminate between SPREAD and MATCH factors, the method can neither prove nor disprove universals.

And of course it can always be the case that MATCH effects don't transpire just because the sample size is too small. Dunn et al.'s sample of four families is definitely too small. It is a truism that each family on earth is subject to specific SPREAD effects (areal histories) that may or may not conflict with MATCH effects. But then, any estimate of the respective contributions of areal vs. universal trends needs a sample of lineages substantially larger than four. A smaller sample may be sufficient for raising hypotheses, but not for testing them, let alone for drawing far-reaching conclusions about the role of society and the brain in language.

### **3. Multivariate modeling and the problem of small families and isolates**

From the argumentation above it follows that what we need in typology is exactly the same approach that is standard in many other disciplines: statistical models of the relative impact of different factors, here SPREAD and MATCH factors. The class of generalized linear models is perfectly suitable for this (Justeson et al. 1990, Bickel 2008, Sinnemäki 2010, Cysouw 2010a), but there are two challenges. One challenge is that at present it is entirely unclear what the most appropriate assumptions are on the underlying stochastic processes (e.g., Poisson or Bernoulli). While certainly important theoretically, this issue has modest practical consequences, typically yielding similar results. The second challenge has wider consequences and is arguably more important: How can we estimate diachronic processes so that we can model the relative impact of MATCH and SPREAD factors on them (rather than on synchronic distributions)?

The proposal by Dunn et al. could provide one response to the second challenge: if applied to a large sample of families, the method would give us estimates of the strength of associations in each family. These estimates could then be taken as the dependent variable in a statistical model based on SPREAD and MATCH factors. This would allow us to estimate the extent to which the factors influence the diachronic tendencies across families and whether these influences are independent of each other. A likely problem is that Dunn et al.'s method requires a relatively large number of datapoints per family and a fairly detailed understanding of family tree structures. But most families in the world are known only through very few representatives – indeed often only a single one: many languages have no known relative and are “isolates”. This risks the possibility that the method cannot incorporate critical data from the majority of languages, calling into question the validity of results.

There are alternative methods that require less datapoints and less knowledge of tree structures and are therefore more in line with the kind of data we have. A crude but still workable method for estimating biases is what I

call the “Family Bias Method”. In this method, each family is evaluated as to whether its daughter languages show a bias towards the same structure (or set of structures), as revealed by a randomization test (Janssen et al. 2006). If there is such a bias (many daughter languages with the same structure), this means that daughter languages have preferentially innovated in the same direction or kept what was already in the proto-language. Either way, a bias suggests that, for whatever reason, there was some systematic preference (and not random fluctuation) in the development of the given family, i.e., a bias suggests a diachronic preference (Bickel 2008).

It is worth emphasizing that in the Family Bias Method persistence of a structure over time is taken as a diachronic signal on a par with signals from structural change: if a certain structure is systematically maintained over time under specific conditions, this suggests a diachronic preference just like systematic change. Or, to put it the other way round: if a factor prefers some structure, this not only means that it will develop easily but also that it will preferably stay once it is developed. In light of this, it is problematic that Dunn et al. (2011) exclude data from Bantu just because the association between structures is invariant. This invariance suggests a very strong preference for diachronic persistence and is therefore an important piece of information.

The Family Bias Method can handle mid-range families, known by about at least half a dozen members. In Bickel 2011 I extend the approach to data from families that we know through less members, or even only through a single member. The basic idea is to use extrapolation algorithms. We first estimate biases in large families (say, with at least five members) and then use this information to estimate the biases that are likely to have been behind the attested structures in small families: if, say, 60 % of large families are biased towards some specific structure (e.g., biased towards OV order, or towards VO order) rather than balanced between structures (i.e., with about as many OV as VO daughters), we estimate a .6 probability that the members of small families come from larger unknown families with a bias as well (in whatever direction), as opposed to families without any bias. Some of the known members will be representative of the bias in the unknown larger family, and so we can take their structural choice (e.g., VO order) to reflect the bias. But some known members will happen to deviate, e.g., the odd guy out that developed an OV pattern although the family as a whole is biased towards VO (like Tunen in Bantu; Mous 1997). The probability of being representative can be estimated from the strength of the bias in large families. For instance, if among biased large families, biases tend to be very strong (e.g., on average covering over 90 % of members), we can estimate a high probability that the known members of small biased families are representative of the larger unknown family from which they derive; then, the probability of being the odd guy(s) out is much lower (though arguably never zero).

In summary, using the probabilities of bias and of representativity based on large families, we can estimate how many of the small families come from larger biased as opposed to unbiased families, and if they come from biased families, we can estimate whether the known members reflect the respective biases of their families or deviate from them. These extrapolation estimates introduce random error but do so along a normal distribution. Therefore, we get a fairly reliable estimate of family biases if we extrapolate many times (say a few thousand times) and compute the average of this.

Estimates on family biases can then be statistically modeled (in a generalized linear model) in a straightforward way. If the estimated biases are best predicted by modeling a *MATCH* factor independently of *SPREAD* factors, this supports the hypothesis that the *MATCH* factor is one of the reasons for the biases. If there is a significant interaction in such a way that the effect of the hypothesized *MATCH* factor systematically differs across the areas defined by the *SPREAD* factor, this suggests non-universality.

#### **4. A case study: VP and NP order**

Here is an application of this to one of the associations examined by Dunn et al. (2011), viz., the correlation between VP and NP orders, more specifically between the relative order of verb and patient (“object”) NPs inside VPs and between head nouns and dependent (“genitive”) NPs inside NPs (Greenberg 1963, Dryer 1992). This correlation is based on a *MATCH* factor of phrase structure harmony that is arguably grounded in processing preferences (e.g., Hawkins 1994, 2004). But since word order is known to have spread in some large areas, we can also expect *SPREAD* factors to influence the distribution NP orders. Map 1 supports this possibility.

In order to examine the relative contribution of these two competing factors, I estimate biases in NP order within families. When families are split between VP orders or areas, I search for the lowest taxon that is not split by these factors and determine family biases within the resulting taxa. In Indo-European and Sino-Tibetan, for example, such taxa can be found at major branch levels. When no lowest non-split taxon can be found or the genealogical database I use (Nichols et al. 2009) does not contain enough subgrouping information, I set up pseudo-taxa based on word order, e.g., an OV pseudo-group and a VO pseudo-group within Tupi-Guaraní. These pseudo-groups are obviously different from true phylogenetic subgroups as established by the kind of arbitrary and idiosyncratic innovations that define genealogical trees. However, since the difference in word order must have resulted from some diachronic change, the pseudo-groups define legitimate sets within which we can examine to what extent they have developed a bias in NP order, given the VP order type and the area. Seeking non-split taxa and pseudo-groups sometimes leaves individ-



Map 1. The distribution of NP order across major areas; ●/○ = Africa, ■/□ = Eurasia, ◆/◇ = New Guinea and Australia, ▲/△ = Americas; black symbols denote GenN, white symbols NGen order (data source: Dryer 2005b)

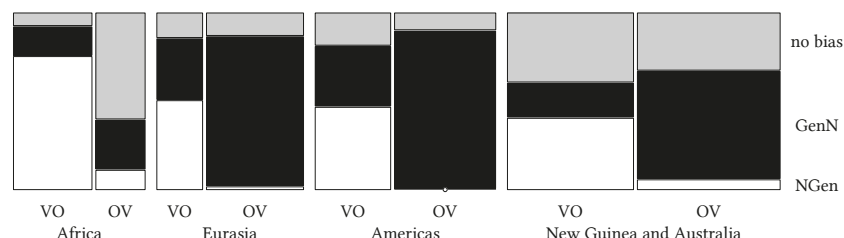


Figure 1. *The distribution of estimated biases in NP order within genealogical taxa across VP order and major areas; the sizes of the individual tiles in the plot are proportional to frequencies, using the “mosaic” plot technique provided by Meyer et al. 2006 (data source: Dryer 2005a, b; Nichols et al. 2009)*

ual languages stranded (e.g., Tunen as the sole OV representative in Bantu, or single-member taxa like Albanian in Indo-European). These are then treated like single-member families such as Basque.<sup>4</sup>

I applied this splitting algorithm to the data provided by Dryer (2005a, b) combined with the genealogical and geographical information from Nichols et al. 2009. These data provide information on 928 languages distributed over 389 non-split taxa. Of these, 45 have five or more members, and this allows tentative estimates of the probabilities of biases and of representativity within each VP order and area. These estimates in turn allow extrapolation to smaller families and isolates. Performing extrapolations 2,000 times and then taking the mean results in what is displayed in Figure 1.

While admittedly tentative, Figure 1 suggests that in all areas considered, OV taxa show a bias towards GenN orders (black tiles) and VO taxa a bias towards NGen orders (white tiles). In addition, there appear to be substantial differences in the extent of this association and the proportion of taxa without any bias (i.e., taxa that mix GenN and NGen orders or are estimated to do so, colored grey in the figure).

In order to explore the relative impact of the VP parameter and the area parameter, I fitted the taxa bias estimates to a series of log-linear models, comparing successively simpler models against more complex models through analyses of deviance (as available in R, cf. R Development Core Team 2011; for the theory see, e.g., Agresti 2002). The counts of taxa estimated as lacking a bias (grey tiles in Figure 1) are excluded from these analyses since unbiased

4. See Bickel 2011 for more detailed exposition and justification of this as well as for a description of the splitting algorithm used here. An R package for estimating family biases is available at <http://www.spw.uzh.ch/software>. The results reported here use the default settings of the package.



taxa contain no evidence for or against a hypothesis: unless we know the proto-language, structural diversity can arise from some languages changing towards the predicted structure (e.g., GenN order in OV taxa) or away from it.

There is no statistical evidence for a three-way interaction ( $LR = 5.56$ ,  $df = 3$ ,  $p = .14$ ).<sup>5</sup> This means that the interactions (statistical associations) of interest, i.e., the interaction between NP and VP order and the interaction between NP order and area, are independent of each other. The interactions of interest themselves are statistically significant, i.e., models without these interactions fit the data significantly less well (interaction between NP and VP order:  $LR = 119.92$ ,  $df = 1$ ,  $p < .001$ ; interaction between NP order and areas:  $LR = 13.83$ ,  $df = 3$ ,  $p = .003$ ).

These results suggest that biases towards specific NP order are associated both with specific VP order and with specific areas. But since the association with VP order is independent of the association with areas, the NP~VP association is universal in the sense that it holds of all regions tested. This supports the hypothesis of a MATCH factor of phrase structure harmony, possibly along the lines proposed by Hawkins (1994, 2004). The fact that Dunn et al.'s method did not pick up any signal from this association is likely to be caused by their extremely small sample of 4 taxa, as opposed to the 389 taxa examined here.

Modeling areas is not at all straightforward since area boundaries are notoriously diffuse and sometimes cross each other. To get a sense of how stable the results are I also fitted a model of areas as proposed by Dryer (1992). This attaches South East Asia to Oceania rather than Eurasia and splits between North and South/Central America. An analysis of deviance of this model suggests a weakly significant three-way interaction, i.e., a model which also includes an interaction of NP order with areas and VP order simultaneously ( $LR = 12.28$ ,  $df = 5$ ,  $p = .031$ ; bootstrap  $p = .046$ ). However, subsequent factorial analysis shows that each area individually contains a significant association of NP and VP order, in the same direction throughout (Africa:  $LR = 11.80$ ,  $p < .001$ ; Eurasia:  $LR = 14.29$ ,  $p < .001$ ; South East Asia and Oceania:  $LR = 36.15$ ,  $p < .001$ ; New Guinea and Australia:  $LR = 7.41$ ,  $p = .007$ ; North America:  $LR = 14.87$ ,  $p < .001$ ; South/Central America:  $LR = 26.99$ ,  $p < .001$ ; all  $df = 1$ ). As a result, the simpler model without the three-way interaction can be taken to fit the data just as well as the more complex model with the three-way interaction. Applying Ockham's razor, this supports the hypothesis of an area-independent NP~VP association.

---

5. *LR*: likelihood ratio, also known as the "deviance" between models. All *p*-values are based on an asymptotic  $\chi^2$ -distribution but are in the same ballpark of what one obtains through resampling (bootstrap) tests (following the recommendation of Janssen et al. 2006, with 2,000 simulations).

As noted before, Figure 1 also suggests substantial variation in the proportion of taxa that are estimated to be without a systematic bias (cf. the grey-colored tiles in the plot). The proportion of biased vs. diverse families is an indicator of the strength of associations: the stronger the biases the heavier the pressure to develop or maintain the preferred structure. For example, in Africa, there seems to have been stronger pressure to develop or maintain a NGen bias in VO groups than a GenN bias in OV groups. To see whether these are significant patterns, I again fitted log-linear models to the data, but now replacing the NP order variable by a “diversity” variable, putting “biased (in either direction)” in binary opposition to “diverse” (i.e., the black and white tiles summed up against the grey tiles in Figure 1).

An analysis of deviance suggests a significant three-way interaction between diversity, VP order, and area ( $LR = 25.83$ ,  $df = 3$ ,  $p < .001$ ). This means that there is no universally constant, area-independent association between having diverse NP orders and the choice of VP order. In Eurasia, the Americas, and in the New Guinea and Australia area, there is no statistical evidence for VP order to affect diversity degrees (all  $p > .10$ ), but there is evidence for such an effect in Africa, where NP orders in OV taxa tend to be more diverse than in VO taxa ( $LR = 23.59$ ,  $df = 1$ ,  $p < .001$ ).<sup>6</sup> Conversely, area appears to be a significant factor in both VP orders (VO:  $LR = 19.66$ ,  $df = 3$ ,  $p < .001$ ; OV:  $LR = 30.19$ ,  $df = 3$ ,  $p < .001$ ). However, post-hoc analyses suggest that this signal comes mostly from the strongly increased proportion of diverse families in the New Guinea and Australia area: the contrast between this area vs. the rest has a highly significant effect on NP order diversity in VO taxa ( $LR = 16.65$ ,  $df = 1$ ,  $p < .001$ ) and a borderline effect in OV taxa ( $LR = 3.06$ ,  $df = 1$ ,  $p = .0804$ ; bootstrap  $p = .081$ ).

These results may be the reason why Dunn et al. (2011) did not pick up a signal for the NP~VP association in the one family they sampled from the New Guinea and Australia area: Austronesian. It is a matter of further research to establish whether in this area there is a general trend across typological variables towards diverse families, perhaps as a result of socially motivated splits and ethnolinguistic compartmentalization. Be this as it may, it is important to recognize that IF families are biased towards some NP order, the direction of the bias is highly correlated with VP order, both here and elsewhere. In other words, what varies is the extent to which there are biases, while their direction reflects universal principles. This is exactly in line with the predictions by Hawkins’s (1994, 2004) theory and by the Greenbergian tradition more generally.

---

6. The fact that an effect transpires in the area with the smallest number of datapoints (cf. the smaller overall tile sizes for Africa in Figure 1) suggests that these findings are not an artifact of small sample sizes.

## 5. Conclusions: What we need

It has long been recognized that the distribution of linguistic structures, and therefore their evolution, is subject to competing motivations. From this it follows that successful modeling and understanding of distributions requires a multivariate approach, e.g., in the form of generalized linear (and other) models. Since at least Dryer 1989 and Nichols 1992 we know that one particularly challenging competitor to universal factors comes from areal histories. The case study presented here shows that areal histories have a significant impact on the type of family biases in NP orders as well as on the degree of these family biases.

With regard to biases in NP order types, however, there is also an association with VP order that is statistically independent of areas. This suggests the presence of a *MATCH* factor, i.e., universality in line with the Greenbergian tradition: no matter where a taxon is located, if it is OV, its members have preferentially developed or maintained GenN order; and if the taxon is a VO group, its members have preferentially developed or maintained NGen order. Independently of this, diachronic preferences for GenN vs. NGen orders also depend on areas: the preference for one or the other order differs from area to area to a statistically significant extent. GenN orders are overall more popular in Eurasia and the Americas, regardless of VP order, and this suggests important *SPREAD* factors inside these areas. The simultaneous relevance of *MATCH* and *SPREAD* factors challenges the century-old opposition between biological vs. social aspects of language (as repeated again by Dunn et al. 2011) and allows for a more realistic understanding of typological distributions.

With regard to degrees of biases (representing strength of diachronic pressures), the association with VP order depends on areas and this suggests area-specific processes rather than universally uniform probabilities of language change (“transition probabilities”, “stability indices”). This confirms the impression that the speed of language change is mostly determined by local sociolinguistic processes, and it opens up research on the areal histories of such processes.

In summary, these results suggest avenues for research beyond simple nature vs. nurture oppositions and for better integration of research on universals with research on areal histories. However, there are several important open issues. First, the Family Bias Method is admittedly crude and we need more refined ways of estimating biases from small families and isolates (e.g., using Bayesian extrapolation methods; Zakharko in preparation). Second, it is unclear whether the Poisson processes assumed by log-linear modeling are fully justified (cf. Cysouw 2010b for recent discussion). Third, areal histories are mostly modeled in terms of categorical distinctions between areas although we know that in reality boundaries are diffuse and have a center vs. periphery structure. It is at

present unclear how this can best be captured when modeling the distribution of family biases.

Thus, there is much important methodological work ahead before we can make bold verdicts on the status of universals in language.

*Received: 31 May 2011*

*Universität Zürich*

*Revised: 21 July 2011*

*Correspondence address:* Seminar für Allgemeine Sprachwissenschaft, Universität Zürich, Plattenstrasse 54, 8032 Zürich, Switzerland; e-mail: [balthasar.bickel@uzh.ch](mailto:balthasar.bickel@uzh.ch)

*Acknowledgements:* The research reported here was supported by Grant Nr. II/83 393 from the Volkswagen Foundation. I would like to thank an anonymous reviewer and Johanna Nichols for helpful comments.

## References

- Agresti, Alan. 2002. *Categorical data analysis*. New York: Wiley-Interscience.
- Bickel, Balthasar. 2008. A general method for the statistical evaluation of typological distributions. Manuscript, Universität Leipzig. [http://www.spw.uzh.ch/bickel/papers/testing\\_universals\\_bickel2008.pdf](http://www.spw.uzh.ch/bickel/papers/testing_universals_bickel2008.pdf)
- Bickel, Balthasar. 2010. Absolute and statistical universals. In Patrick Colm Hogan (ed.), *The Cambridge encyclopedia of the language sciences*, 77–79. Cambridge: Cambridge University Press.
- Bickel, Balthasar. 2011. Distributional biases in language families. Manuscript, Universität Leipzig. <http://www.spw.uzh.ch/bickel/papers/stability.fsjn.2011bickelrevised.pdf>
- Curie, Thomas E., Simon J. Greenhill & Ruth Mace. 2009. Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philosophical Transactions of the Royal Society B* 365. 3909–3912.
- Cysouw, Michael. 2002. Interpreting typological clusters. *Linguistic Typology* 6. 49–93.
- Cysouw, Michael. 2010a. Dealing with diversity: Towards an explanation of NP-internal word order frequencies. *Linguistic Typology* 14. 253–286.
- Cysouw, Michael. 2010b. On the probability distribution of typological frequencies. In Christian Ebert, Gerhard Jäger & Jens Michaelis (eds.), *The mathematics of language*, 29–35. Berlin: Springer.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–13. Cambridge, MA: MIT Press.
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
- Dryer, Matthew S. 2005a. Order of subject, object, and verb. In Haspelmath et al. (eds.) 2005, 330–341.
- Dryer, Matthew S. 2005b. Order of genitive and noun. In Haspelmath et al. (eds.) 2005, 350–353.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.
- Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20. 271–287.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Janssen, Dirk, Balthasar Bickel & Fernando Zúñiga. 2006. Randomization tests in language typology. *Linguistic Typology* 10. 419–440.
- Justeson, John S. & Laurence D. Stephens. 1990. Explanation for word order universals: A log-linear analysis. In Werner Bahner, Joachim Schild & Dieter Viehweger (eds.), *Proceedings of the XIV International Congress of Linguists*, Vol. 3, 2372–2376. Berlin: Akademie-Verlag.
- Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4. 307–333.
- Meyer, David, Achim Zeileis & Kurt Hornik. 2006. The strucplot framework: Visualizing multi-way contingency tables with vcd. *Journal of Statistical Software* 17. 1–48.
- Mous, Maarten. 1997. The position of the object in Tunen. In Rose-Marie Déchaine & Victor Manfredi (eds.), *Object positions in Benue-Kwa*, 123–137. Den Haag: Holland Academic Graphics.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.
- Nichols, Johanna & Balthasar Bickel. 2009. The AUTOTYP genealogy and geography database: 2009 release. Electronic database. <http://www.spw.uzh.ch/autotyp>
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. <http://www.r-project.org>
- Sinnemäki, Kaius. 2010. Word order in zero-marking languages. *Studies in Language* 34. 869–912.
- Zakharko, Taras (in preparation). A Bayesian approach to estimating typological biases in language families. Manuscript, Universität Zürich.